Lessons we've learned: Tips and techniques for online data collection.

Tom Buchanan

Department of Psychology,
University of Westminster, UK.

buchant@wmin.ac.uk    www.buchanan.org.uk
Online Data Collection: What, Why, How?

• Collection of psychological test data via Internet (usually WWW forms), for research or applied purposes

• Many advantages – sample size, cost, efficiency…

• Often forms hosted on survey / research sites or personal pages; constructed with authoring systems or hand-coded html
Opportunities and Threats

- Sampling (numbers, rare populations, ease of recruitment)
- Efficiency
- Cost
- Increased self-disclosure
- Metadata available (latencies, dropout curves)
- Advanced techniques possible
- Ethical advantages

- Sampling (biases, voluntariness, recruitment can be tough)
- Loss of control (e.g. environmental variance, technical issues).
- Inauthentic / multiple responding
- Interactions between sample / topic / method.
- Ethical pitfalls
Key Considerations

• Validity of measures and measurement invariance: Quantitative and qualitative equivalence across formats.

• Authenticity of responding.

• Biases (not just sampling).
Validity and Equivalence: Sample Findings

• Buchanan et al (2005a) – validated online IPIP Five Factor inventory – has expected latent structure and expected associations with other variables.
• Buchanan et al (2005b) – latent structure of online Prospective Memory Questionnaire differed markedly from published (offline) structure (qualitative non-equivalence).
• Buchanan (2003) found online Hospital Anxiety and Depression Scale scores very high in comparison with offline patient norms (qualitative but not quantitative equivalence).

• Along with other studies, implies measurement invariance possible but not guaranteed
Authenticity: Verification of Self-Reports

- Scholey et al, 2011: \( n=49 \) did online drug questionnaire (in lab), gave hair samples.
- Self-reported MDMA use associated with detection in hair sample (\( X^2(2)=20.8689, p=0.000032 \)).

<table>
<thead>
<tr>
<th>MDMA</th>
<th>Never Used</th>
<th>Used 1-9 times</th>
<th>Used 10-100 times</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detected</td>
<td>1</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>Not detected</td>
<td>19</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

- Supports authenticity of online self-reports.
Bias? Personality, Executive Function and Digit Span

• $N=354$ completed online Five Factor inventory, self-report measure of executive problems, digit span task (forward and reverse).

• Expected executive problems to correlate negatively with reverse digit span.
  – Buchanan et al (2010) found $r=-.42$ between online (lab) exec questions and offline reverse digit span.
(Preliminary) Findings

• Executive questions did not correlate with reverse digit span ($r= -0.08$, $n=354$, $p > 0.05$).
• Reverse digit span did correlate with personality variables – Openness to Experience ($r=0.159$, $n=354$, $p < 0.005$) and Agreeableness ($r=0.134$, $n=354$, $p < 0.05$).
• Why? ‘experimenter’ / motivation effects?
• Potential issue: non-cognitive factors may affect online cognitive assessments in unexpected ways.
Unanswered Questions

• Under what circumstances may online and offline assessments differ?
• Biases and confounding effects – need to document and understand.
• Are online measures really ‘better’ – candour, authenticity, ‘Real Me’ hypotheses?
• Use and effects of ‘amateur’ tests
• Use of normative data – what norms are appropriate?
• How much do these kind of issues matter?
Suggestions for Good Practice

- Build on existing work, don’t reinvent wheel
- Keep it simple – use ubiquitous technologies, and off-the-shelf solutions where possible
- Be wary of biases and confounds
- Answer those unanswered questions!
References